Redefine statistical significance

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, David Cesarini, Christopher D. Chambers, Merlise Clyde, Thomas D. Cook, Paul De Boeck, Zoltan Dienes, Anna Dreber, Kenny Easwaran, Charles Efferson, Ernst Fehr, Fiona Fidler, Andy P. Field, Malcolm Forster, Edward I. George, Richard Gonzalez, Steven Goodman, Edwin Green, Donald P. Green, Anthony Greenwald, Jarrod D. Hadfield, Larry V. Hedges, Leonhard Held, Teck Hua Ho, Herbert Hoijtink, Daniel J. Hruschka, Kosuke Imai, Guido Imbens, John P. A. Ioannidis, Minjeong Jeon, James Holland Jones, Michael Kirchler, David Laibson, John List, Roderick Little, Arthur Lupia, Edouard Machery, Scott E. Maxwell, Michael McCarthy, Don Moore, Stephen L. Morgan, Marcus Munafó, Shinichi Nakagawa, Brendan Nyhan, Timothy H. Parker, Luis Pericchi, Marco Perugini, Jeff Rouder, Judith Rousseau, Victoria Savalei, Felix D. Schönbrodt, Thomas Sellke, Betsy Sinclair, Dustin Tingley, Trisha Van Zandt, Simine Vazire, Duncan J. Watts, Christopher Winship, Robert L. Wolpert, Yu Xie, Cristobal Young, Jonathan Zinman and Valen E. Johnson

Daniel J. Benjamin[1*], James O. Berger[2], Magnus Johannesson[3*], Brian A. Nosek[4,5], E.-J. Wagenmakers[6], Richard Berk[7,10], Kenneth A. Bollen[8], Björn Brembs[9], Lawrence Brown[10], Colin Camerer[11], David Cesarini[12,13], Christopher D. Chambers[14], Merlise Clyde[2], Thomas D. Cook[15,16], Paul De Boeck[17], Zoltan Dienes[18], Anna Dreber[3], Kenny Easwaran[19], Charles Efferson[20], Ernst Fehr[21], Fiona Fidler[22], Andy P. Field[18], Malcolm Forster[23], Edward I. George[10], Richard Gonzalez[24], Steven Goodman[25], Edwin Green[26], Donald P. Green[27], Anthony G. Greenwald[28], Jarrod D. Hadfield[29], Larry V. Hedges[30], Leonhard Held[31], Teck Hua Ho[32], Herbert Hoijtink[33], Daniel J. Hruschka[34], Kosuke Imai[35], Guido Imbens[36], John P. A. Ioannidis[37], Minjeong Jeon[38], James Holland Jones[39,40], Michael Kirchler[41], David Laibson[42], John List[43], Roderick Little[44], Arthur Lupia[45], Edouard Machery[46], Scott E. Maxwell[47], Michael McCarthy[48], Don A. Moore[49], Stephen L. Morgan[50], Marcus Munafó[51,52], Shinichi Nakagawa[53], Brendan Nyhan[54], Timothy H. Parker[55], Luis Pericchi[56], Marco Perugini[57], Jeff Rouder[58], Judith Rousseau[59], Victoria Savalei[60], Felix D. Schönbrodt[61], Thomas Sellke[62], Betsy Sinclair[63], Dustin Tingley[64], Trisha Van Zandt[65], Simine Vazire[66], Duncan J. Watts[67], Christopher Winship[68], Robert L. Wolpert[2], Yu Xie[69], Cristobal Young[70], Jonathan Zinman[71] and Valen E. Johnson[72*]

[1]Center for Economic and Social Research and Department of Economics, University of Southern California, Los Angeles, CA 90089-3332, USA. [2]Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA. [3]Department of Economics, Stockholm School of Economics, Stockholm SE-113 83, Sweden. [4]University of Virginia, Charlottesville, VA 22908, USA. [5]Center for Open Science, Charlottesville, VA 22903, USA. [6]Department of Psychology, University of Amsterdam, Amsterdam 1018 VZ, The Netherlands. [7]School of Arts and Sciences and Department of Criminology, University of Pennsylvania, Philadelphia, PA 19104-6286, USA. [8]Department of Psychology and Neuroscience, Department of Sociology, University of North Carolina Chapel Hill, Chapel Hill, NC 27599-3270, USA. [9]Institute of Zoology — Neurogenetics, Universität Regensburg, Universitätsstrasse 31, 93040 Regensburg, Germany. [10]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. [11]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [12]Department of Economics, New York University, New York, NY 10012, USA. [13]The Research Institute of Industrial Economics (IFN), Stockholm SE-102 15, Sweden. [14]Cardiff University Brain Research Imaging Centre (CUBRIC), Cardiff CF24 4HQ, UK. [15]Northwestern University, Evanston, IL 60208, USA. [16]Mathematica Policy Research, Washington, DC 20002-4221, USA. [17]Department of Psychology, Quantitative Program, Ohio State University, Columbus, OH 43210, USA. [18]School of Psychology, University of Sussex, Brighton BN1 9QH, UK. [19]Department of Philosophy, Texas A&M University, College Station, TX 77843-4237, USA. [20]Department of Psychology, Royal Holloway University of London, Egham Surrey TW20 0EX, UK. [21]Department of Economics, University of Zurich, 8006 Zurich, Switzerland. [22]School of BioSciences and School of Historical & Philosophical Studies, University of Melbourne, Parkville, VIC 3010, Australia. [23]Department of Philosophy, University of Wisconsin — Madison, Madison, WI 53706, USA. [24]Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1043, USA. [25]Stanford University, General Medical Disciplines, Stanford, CA 94305, USA. [26]Department of Ecology, Evolution and Natural Resources SEBS, Rutgers University, New Brunswick, NJ 08901-8551, USA. [27]Department of Political Science, Columbia University in the City of New York, New York, NY 10027, USA. [28]Department of Psychology, University of Washington, Seattle, WA 98195-1525, USA. [29]Institute of Evolutionary Biology School of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3JT, UK. [30]Weinberg College of Arts & Sciences Department of Statistics, Northwestern University, Evanston, IL 60208, USA. [31]Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich, 8001 Zurich, Switzerland. [32]National University of Singapore, Singapore 119077, Singapore. [33]Department of Methods and Statistics, Universiteit Utrecht, Utrecht 3584 CH, The Netherlands. [34]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287-2402, USA. [35]Department of Politics and Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08544, USA. [36]Stanford University, Stanford, CA 94305-5015, USA. [37]Departments of Medicine, of Health Research and Policy, of Biomedical Data Science, and of Statistics and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA 94305, USA. [38]Advanced Quantitative Methods, Social Research Methodology, Department of Education, Graduate School of Education & Information Studies, University of California, Los Angeles, CA 90095-1521, USA. [39]Department of Life Sciences, Imperial College London, Ascot

Supplementary Information:

Supplementary Text

R code used to generate Figures 1 and 2

Supplementary Information:

Supplementary Text

Figure 1

All four curves in Figure 1 describe the relationship between (i) a P-value based on a two-sided normal test and (ii) a Bayes factor or a bound on a Bayes factor. The P-values are based on a two-sided test that the mean μ of an independent and identically distributed sample of normally distributed random variables is 0. The variance of the observations is known. Without loss of generality, we assume that the variance is 1 and the sample size is also 1. The curves in the figure differ according to the alternative hypotheses that they assume for calculating (ii).

Because these curves involve two-sided tests, all alternative hypotheses are restricted to be symmetric around 0. That is, the density assumed for the value of μ under the alternative hypothesis is always assumed to satisfy $f(\mu) = f(-\mu)$.

The curve labeled "Power" corresponds to defining the alternative hypothesis so that power is 75% in a two-sided 5% test. This is achieved by assuming that μ under the alternative hypothesis is equal to $\pm(z_{0.025} + z_{0.75}) = \pm 2.63$. That is, the alternative hypothesis places ½ its prior mass on 2.63 and ½ its mass on -2.63.

The curve labeled UMPBT corresponds to the uniformly most powerful Bayesian test (*2*) that corresponds to a classical, two-sided test of size $\alpha = 0.005$. The alternative hypothesis for this Bayesian test places ½ mass at 2.81 and ½ mass at -2.81. The null hypothesis for this test is rejected if the Bayes factor exceeds 25.7. Note that this curve is nearly identical to the "Power" curve if that curve had been defined using 80% power, rather than 75% power. The Power curve for 80% power would place ½ its mass at ± 2.80.

The Likelihood Ratio Bound curve represents an approximate upper bound on the Bayes factor obtained by defining the alternative hypothesis as putting ½ its mass on $\pm\bar{x}$, where $\bar{x}$ is the observed sample mean. Over the range of P-values displayed in the figure, this alternative hypothesis very closely approximates the maximum Bayes factor that can be attained from among the set of alternative hypotheses constrained to be of the form $0.5 \times [f(\mu) + f(-\mu)]$ for some density function f.

The Local-H_1 curve is described fully in the figure caption. A fuller explanation and discussion of this bound can be found in ref. *15*.

Equation 2 and Figure 2

This equation defines the large-sample relationship between the false positive rate, power $1 - \beta$, type I error rate α, and the probability that the null hypothesis is true when a large number of independent experiments have been conducted. More specifically, suppose that n independent hypothesis tests are conducted, and suppose that in each test the probability that the null hypothesis is true is ϕ. If the null hypothesis is true, assume that the probability that it is falsely rejected (i.e., a false positive occurs) is α. For the test $j = 1, \dots, n$, define the random variable $X_j = 1$ if the null hypothesis is true *and* the null hypothesis is rejected, and $X_j = 0$ if either the alternative hypothesis is true or the null hypothesis is not rejected. Note that the X_j are independent Bernoulli random variables with $\Pr(X_j = 1) = \alpha\phi$. Also for test *j*, define another random variable $Y_j = 1$ if the alternative hypothesis is true *and* the null hypothesis is rejected, and 0 otherwise. It follows that the Y_j are independent Bernoulli random variables with $\Pr(Y_j = 1) = (1-\phi)(1-\beta)$. Note that Y_j is independent of Y_k for $j \neq k$, but Y_j is not independent of X_j. For the *n* experiments, the false positive rate can then be written as:

$$FPR = \frac{\sum_{j=1}^{n} X_j}{\sum_{j=1}^{n} X_j + \sum_{j=1}^{n} Y_j} = \frac{\sum_{j=1}^{n} X_j/n}{\sum_{j=1}^{n} X_j/n + \sum_{j=1}^{n} Y_j/n}.$$

By the strong law of large numbers, $\sum_{j=1}^{n} X_j/n$ converges almost surely to $\alpha\phi$, and $\sum_{j=1}^{n} Y_j/n$ converges almost surely to $(1-\phi)(1-\beta)$. Application of the continuous mapping theorem yields

$$FPR \xrightarrow{\text{a.s.}} \frac{\alpha\phi}{\alpha\phi + (1-\phi)(1-\beta)}.$$

Figure 2 illustrates this relationship for various values of α and prior odds for the alternative, $\frac{1-\phi}{\phi}$.

R code used to generate Figure 1:

```
type1=.005
type1Power=0.05
type2=0.25
p=1-c(9000:9990)/10000
xbar = qnorm(1-p/2)

# alternative based on 80% POWER IN 5% TEST
muPower = qnorm(1-type2)+qnorm(1-type1Power/2)
bfPow = 0.5*(dnorm(xbar,muPower,1)+dnorm(xbar,-muPower,1))/dnorm(xbar,0,1)

muUMPBT = qnorm(0.9975)
bfUMPBT = 0.5*(dnorm(xbar,muUMPBT,1)+dnorm(xbar,-
muUMPBT,1))/dnorm(xbar,0,1)

# two-sided "LR" bound
bfLR = 0.5/exp(-0.5*xbar^2)

bfLocal = -1/(2.71*p*log(p))

#coordinates for dashed lines
data = data.frame(p,bfLocal,bfLR,bfPow,bfUMPBT)
U_005 = max(data$bfLR[data$p=="0.005"])
L_005 = min(data$bfLocal[data$p=="0.005"])
U_05 = max(data$bfLR[data$p=="0.05"])
L_05 = min(data$bfUMPBT[data$p=="0.05"])

# Local bound; no need for two-sided adjustment


#plot margins
par(mai=c(0.8,0.8,.1,0.4))
par(mgp=c(2,1,0))

matplot(p,cbind(bfLR,-1/(2.71*p*log(p))),type='n',log='xy',
        xlab=expression(paste(italic(P) ,"-value")),
        ylab="Bayes Factor",
        ylim = c(0.3,100),
        bty="n",xaxt="n",yaxt="n")
lines(p,bfPow,col="red",lwd=2.5)
lines(p,bfLR,col="black",lwd=2.5)
lines(p,bfUMPBT,col="blue",lwd=2.5)
lines(p,bfLocal,col="green",lwd=2.5)
legend(0.015,100,c(expression(paste("Power")),"Likelihood Ratio
Bound","UMPBT",expression(paste("Local-",italic(H)[1],"
Bound"))),lty=c(1,1,1,1),
       lwd=c(2.5,2.5,2.5,2.5),col=c("red","black","blue","green"), cex =
0.8)
#text(0.062,65, "\u03B1", font =3, cex = 0.9)

#customizing axes
#x axis
axis(side=1,at=c(-2,0.001,0.0025,0.005,0.010,0.025,0.050,0.100,0.14),
```

```
     labels =
c("","0.0010","0.0025","0.0050","0.0100","0.0250","0.0500","0.1000",""),lw
d=1,
     tck = -0.01, padj = -1.1, cex.axis = .8)
#y axis on the left - main
axis(side=2,at=c(-0.2, 0.3,0.5,1,2,5,10,20,50,100),labels =
c("","0.3","0.5","1.0","2.0","5.0","10.0","20.0","50.0","100.0"),lwd=1,las
= 1,
     tck = -0.01, hadj = 0.6, cex.axis = .8)
#y axis on the left - secondary (red labels)
axis(side=2,at=c(L_005,U_005),labels = c(13.9,25.7),lwd=1,las= 1,
     tck = -0.01, hadj = 0.6, cex.axis = .6,col.axis="red")
#y axis on the right - main
axis(side=4,at=c(-0.2, 0.3,0.5,1,2,5,10,20,50,100),labels =
c("","0.3","0.5","1.0","2.0","5.0","10.0","20.0","50.0","100.0"),lwd=1,las
= 1,
     tck = -0.01, hadj = 0.4, cex.axis = .8)
#y axis on the right - secondary (red labels)
axis(side=4,at=c(L_05,U_05),labels = c(2.4,3.4),lwd=1,las= 1,
     tck = -0.01, hadj = 0.4, cex.axis = .6,col.axis="red")

###dashed lines
segments(x0 = 0.000011, y0= U_005, x1 = 0.005, y1 = U_005, col = "gray40",
lty = 2)
segments(x0 = 0.000011, y0= L_005, x1 = 0.005, y1 = L_005, col = "gray40",
lty = 2)
segments(x0 = 0.005, y0= 0.00000001, x1 = 0.005, y1 = U_005, col =
"gray40", lty = 2)

segments(x0 = 0.05, y0= U_05, x1 = 0.14, y1 = U_05, col = "gray40", lty =
2)
segments(x0 = 0.05, y0= L_05, x1 = 0.14, y1 = L_05, col = "gray40", lty =
2)
segments(x0 = 0.05, y0= 0.00000001, x1 = 0.05, y1 = U_05, col = "gray40",
lty = 2)
```

R code used to generate Figure 2:

```
pow1=c(5:999)/1000   # power range for 0.005 tests
pow2=c(50:999)/1000  # power range for 0.05 tests
alpha=0.005 # test size
pi0=5/6  # prior probability
N=10^6  # doesn't matter


#graph margins
par(mai=c(0.8,0.8,0.1,0.1))
par(mgp=c(2,1,0))


plot(pow1,alpha*N*pi0/(alpha*N*pi0+pow1*(1-pi0)*N),type='n',ylim = c(0,1),
xlim = c(0,1.5),
     xlab='Power                                      ',
     ylab='False positive rate', bty="n", xaxt="n", yaxt="n")
#grid lines
segments(x0 = -0.058, y0 = 0, x1 = 1, y1 = 0,lty=1,col = "gray92")
segments(x0 = -0.058, y0 = 0.2, x1 = 1, y1 = 0.2,lty=1,col = "gray92")
segments(x0 = -0.058, y0 = 0.4, x1 = 1, y1 = 0.4,lty=1,col = "gray92")
segments(x0 = -0.058, y0 = 0.6, x1 = 1, y1 = 0.6,lty=1,col = "gray92")
segments(x0 = -0.058, y0 = 0.8, x1 = 1, y1 = 0.8,lty=1,col = "gray92")
segments(x0 = -0.058, y0 = 1, x1 = 1, y1 = 1,lty=1,col = "gray92")


lines(pow1,alpha*N*pi0/(alpha*N*pi0+pow1*(1-
pi0)*N),lty=1,col="blue",lwd=2)
odd_1_5_1 = alpha*N*pi0/(alpha*N*pi0+pow1[995]*(1-pi0)*N)
alpha=0.05
pi0=5/6
lines(pow2,alpha*N*pi0/(alpha*N*pi0+pow2*(1-
pi0)*N),lty=2,col="blue",lwd=2)
odd_1_5_2 = alpha*N*pi0/(alpha*N*pi0+pow2[950]*(1-pi0)*N)


alpha=0.05
pi0=10/11
lines(pow2,alpha*N*pi0/(alpha*N*pi0+pow2*(1-pi0)*N),lty=2,col="red",lwd=2)
odd_1_10_2 = alpha*N*pi0/(alpha*N*pi0+pow2[950]*(1-pi0)*N)
alpha=0.005
pi0=10/11
lines(pow1,alpha*N*pi0/(alpha*N*pi0+pow1*(1-pi0)*N),lty=1,col="red",lwd=2)
odd_1_10_1 = alpha*N*pi0/(alpha*N*pi0+pow1[995]*(1-pi0)*N)

alpha=0.05
pi0=40/41
lines(pow2,alpha*N*pi0/(alpha*N*pi0+pow2*(1-
pi0)*N),lty=2,col="green",lwd=2)
odd_1_40_2 = alpha*N*pi0/(alpha*N*pi0+pow2[950]*(1-pi0)*N)
alpha=0.005
pi0=40/41
```

```
lines(pow1,alpha*N*pi0/(alpha*N*pi0+pow1*(1-
pi0)*N),lty=1,col="green",lwd=2)
odd_1_40_1 = alpha*N*pi0/(alpha*N*pi0+pow1[995]*(1-pi0)*N)




#customizing axes
axis(side=2,at=c(-0.5,0,0.2,0.4,0.6,0.8,1.0),labels =
c("","0.0","0.2","0.4","0.6","0.8","1.0"),
     lwd=1,las= 1,tck = -0.01, hadj = 0.4, cex.axis = .8)
axis(side=1,at=c(-0.5,0,0.2,0.4,0.6,0.8,1.0),labels =
c("","0.0","0.2","0.4","0.6","0.8","1.0"),
     lwd=1,las= 1, tck = -0.01, padj = -1.1, cex.axis = .8)



legend(1.05,1,c("Prior odds = 1:40","Prior odds = 1:10","Prior odds =
1:5"),pch=c(15,15,15),
       col=c("green","red","blue"), cex = 1)


###############  Use these commands to add brackets in Figure 2

library(pBrackets)


#add text and brackets
text(1.11,(odd_1_5_2+odd_1_40_2)/2, expression(paste(italic(P)," < 0.05
threshold")), cex = 0.9,adj=0)
text(1.11,(odd_1_5_1+odd_1_40_1)/2, expression(paste(italic(P)," < 0.005
threshold")), cex = 0.9,adj=0)
brackets(1.03, odd_1_40_1, 1.03, odd_1_5_1, h = NULL, ticks = 0.5,
curvature = 0.7, type = 1,
        col = 1, lwd = 1, lty = 1, xpd = FALSE)
brackets(1.03, odd_1_40_2, 1.03, odd_1_5_2, h = NULL, ticks = 0.5,
curvature = 0.7, type = 1,
         col = 1, lwd = 1, lty = 1, xpd = FALSE)
```